The Journal of Child Psychology and Psychiatry

The Association for Child and Adolescent Mental Health

Journal of Child Psychology and Psychiatry 64:5 (2023), pp 834–838 doi:10.1111/jcpp.13802

Commentary: Best practices and *processes* for assessment of autism spectrum disorder – the intended role of standardized diagnostic instruments

Somer L. Bishop,[1] **and Catherine Lord**[2]

[1]Department of Psychiatry and Behavioral Sciences and Weill Institute for Neurosciences, University of California, San Francisco, San Francisco, CA, USA; [2]Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, Los Angeles, CA, USA

Given the relatively low recognized prevalence of autism prior to 2000, and the extreme phenotypic heterogeneity of the condition, researchers have long recognized the value of combining samples from different sites or replicating study results across modestly sized samples. However, because autism is a behaviorally defined condition, diagnostic practices may vary widely between practitioners, prompting concerns about how autism cases are defined. Even before autism was formally recognized as a DSM diagnosis, major efforts were undertaken in the 1960s and 1970s to establish standard diagnostic criteria to facilitate the accurate identification of patients for epidemiological or other larger-scale studies (Evans, 2013; Lotter, 1966; Rutter, 1966). Later, in the 1990s–2000s, the development of standardized diagnostic and screening practices became a focus of clinical research (Gillberg, 2013; Lord, Elsabbagh, Baird, & Veenstra-Vanderweele, 2018; Rosen, Lord, & Volkmar, 2021).

As the pace of autism research accelerated, interest in characterizing participants using similar methods propelled widespread adoption of standardized instruments such as the Autism Diagnostic Interview (ADI; Le Couteur, 1989)/Autism Diagnostic Interview-Revised (ADI-R; Rutter, Le Couteur, & Lord, 2003), and the Pre-linguistic Autism Diagnostic Observation Schedule (PLADOS; DiLavore, Lord, & Rutter, 1995)/Autism Diagnostic Observation Schedule (ADOS; Lord, Rutter, DiLavore, & Risi, 2001)/Autism Diagnostic Observation Schedule-2 (ADOS-2; Lord et al., 2012) in research protocols. The original intent of these tools was to formalize the procedures through which clinicians and researchers gather information about autism-related symptoms in patients referred due to concern for autism spectrum disorder (ASD), thus providing standardized methods for eliciting descriptions or observations of behaviors that are diagnostically relevant. Moreover, recognizing that the *process* of administering and scoring these tools is at least as valuable as the scores they yield, the ADI-R and ADOS were designed for use *by* diagnosticians. This differs significantly from labs or other medical tests that can be "ordered," with results later interpreted by the diagnostician. With a behaviorally defined condition, such as ASD, where conceptualizations of what is or is not characteristic of the condition change over time, and exclusive reliance on self-report measures is not possible—the administration of diagnostic tests simply cannot be divorced from the clinician's expertise and understanding of that condition.

While standardized ASD diagnostic instruments were designed to be used by individuals well-versed in ASD diagnosis, even experts benefit from systematic procedures for collecting information. Studies show that quick diagnostic judgments by expert clinicians lead to high rates of false positives and false negatives (Gabrielsen et al., 2015). Importantly, the contexts that we create for hearing about or directly observing behaviors may drastically affect the information that we obtain (Lord, Rutter, & DiLavore, 1999). Thus, when clinicians are idiosyncratic or overly hasty in their information-gathering efforts, the potential for error and bias increases. In the observational context, people behave differently in a room full of toys than in a room with only furniture. Children behave differently when adults attempt to engage them directly in play or conversation, as opposed to when they are left to play on their own. People also behave differently when faced with tasks that are developmentally appropriate versus those that are too advanced or too juvenile. Therefore, "The goal of the ADOS-G is to provide presses that elicit spontaneous behaviors in standardized contexts. Structured activities and materials, and less structured interactions, provide standard contexts within the ADOS-G in which social, communicative, and other behaviors relevant to the understanding of PDDs are observed" (Lord et al., 1999, p. 205). In addition, reports from parents and caregivers about the developmental history and pervasiveness of ASD-related behaviors are also essential to the diagnostic assessment. The ADOS provides a relatively brief observation, during which information about certain behaviors, such as restricted, repetitive, and sensory interests or behaviors (RRBs), or "developing and maintaining

---

Conflict of interest statement: See Acknowledgements for full disclosures.

relationships," may be difficult to elicit. Further, RRBs significantly decrease with age (Bishop, Richler, & Lord, 2006; Esbensen, Seltzer, Lam, & Bodfish, 2009; Uljarević et al., 2022) and may manifest differently in some girls with ASD (Kaat et al., 2020). Therefore, relying on families to spontaneously report behaviors that they consider repetitive, as opposed to systematically collecting a detailed history of behaviors that a clinician could appropriately conceptualize as examples of RRBs, risks missing diagnostically critical symptoms in certain patients.

Although standardized instruments offer several potential advantages in the diagnostic assessment of autism, *which* specific tools can or should be used vary according to several factors. For example, the ADI-R and ADOS were not developed for individuals with severe vision, hearing, and/or motor impairments, nor are they valid for individuals with profound intellectual disabilities (Lord, Luyster, Gotham, & Guthrie, 2012). Because individuals with these conditions make up a minority of individuals referred for ASD diagnostic assessment and require additional specialized assessment procedures (Thurm et al., 2021), they were not represented in the original validation samples. However, compared to the general population, rates of autism are higher in blind and deaf individuals and those with severe/profound intellectual disabilities (Rosenhall, Nordin, Sandstrom, Ahlsen, & Gillberg, 1999; Rydzewska et al., 2018). Thus, to say that they cannot be diagnosed with autism is both inaccurate and clinically irresponsible.

Special circumstances, such as the examples above, require modifications to typical assessment procedures. Adaptations of existing instruments for use in special populations are beginning to appear, but we must be careful about implementing new tools or procedures that have been developed based on anecdotal observations of a particular group (e.g., females), and that may only reflect the opinions of one or a few clinicians. Instead, the field will benefit from the development of empirically derived tools, based on careful and systematic research, which are more appropriately targeted and have improved diagnostic validity for certain groups of patients (Bal et al., 2020; Phillips et al., 2022; Wright et al., 2022). Modifications to typical assessment practices are also required when a child or their family do not speak the same language as the diagnostician, as standardized instruments cannot be validly administered via a translator, or when there is no living parent/caregiver who is familiar enough with the individual's developmental history to complete an interview like the ADI-R. However, the overall *goals* of the assessment—to collect information needed to make diagnoses and inform treatment planning—remain the same. The importance of flexibility in ASD diagnostic procedures became a focus of several papers during the COVID-19 pandemic when in-person testing was prohibited or severely restricted (e.g., requiring PPE; Spain et al., 2022; Wagner et al., 2021; Zwaigenbaum et al., 2021). During this time, clinicians adapted the use of existing tools and developing new tools (Dow et al., 2022), so that diagnostic assessments could continue, thus providing another example of how clinicians must modify specific assessment procedures depending on the circumstances.

Here we come to the critical point: best practice *processes* for diagnostic assessment exist independently of the use of any single tool. Standardized diagnostic instruments were developed to aid the expert, to structure the assessment to ensure that diagnostically relevant information would be available for making diagnostic judgments. Never were they meant to prevent access to appropriate services. And yet, healthcare delivery and funding systems are increasingly trying to mandate the use of specific tools in the diagnostic assessment of ASD. This can be extremely damaging in situations when standardized instruments cannot be validly administered. In addition, blanket requirements directly contradict best practice recommendations for individualizing assessment procedures.

A related problem that comes about with mandating the use of specific tools is when systems require that an individual's scores exceed ASD "cutoffs" in order to access appropriate clinical services. As indicated above, an original goal of standardized ASD diagnostic instruments was to provide researchers with a way of more validly classifying ASD "cases." Instrument thresholds for yes/no ASD classification were identified to maximize diagnostic validity—as measured by agreement with best-estimate clinical diagnoses of ASD. However, such thresholds are never capable of yielding perfect sensitivity or specificity values, and they are highly dependent on the samples from which they were derived. This means that there will always be individuals with ASD whose scores fail to meet cutoffs, as well as individuals without ASD whose scores incorrectly exceed cutoffs. As with any tool, clinicians must be empowered to make an overall best-estimate clinical diagnosis, using the process of standardized test administration, regardless of instrument-derived ASD/non-ASD classification. This is especially relevant in light of increasingly complex clinical referral populations, comprised of individuals with ASD with less prototypical ASD symptom presentations, as well as individuals without ASD with various psychiatric and medical diagnoses, all of whom may be likely to receive scores very near (and on either side of) the established cutoffs (Elias & Lord, 2022; Klaiman et al., 2022; Lord & Bishop, 2021; McDonnell et al., 2019).

Ironically, attempts to require the administration of standardized instruments like the ADOS-2 are likely rooted in a shared goal of enhancing the validity and reliability of diagnoses. For example, requiring the ADOS was probably intended to serve as a proxy for whether the assessment included a direct observation, which is central to best practice. Relying exclusively on parent-report measures is likely to result in a high number of false positives (and false negatives) (Havdahl et al., 2016; Hus, Bishop, Gotham, Huerta, & Lord, 2013), meaning that systems could risk serving and paying for large numbers of children who did not have autism if the direct observation was not required. Further, when the ADOS was initially developed, it was administered by people with extensive training in autism. This is reflected in the validity data, which were collected by experienced and research reliable examiners, and which show high levels of sensitivity and specificity using expert clinical diagnosis as the reference gold standard. Therefore, it was likely hoped that requiring an ADOS would ensure that the diagnosis was made by an expert.

Sadly, the shared goal of enhancing diagnostic assessment practices by encouraging the use of standardized instruments appears to have been lost in translation. Systems have come to focus too much on scores from certain tools, and too little on what can be gained from the actual administration and scoring of those tools. Moreover, requirements that certain tools, such as the ADOS, be used in diagnostic assessment have resulted in large groups of professionals who are technically trained on the ADOS (e.g., after completing a 2-day introductory training), but who lack broader training in assessment and differential diagnosis of autism and other neurodevelopmental disorders. Such professionals may be inappropriately emboldened to diagnose or rule out autism because of the overemphasis on test results, rather than assessment processes. This is especially troublesome given that ADOS scores obtained from administrations by minimally trained and/or nonresearch-reliable examiners are highly unlikely to work as well as the validity data would suggest. On the other hand, attempts to efficiently train large numbers of providers on autism diagnostic instruments result directly from massive shortages of expert clinicians, and numbers of assessment referrals that far exceed any clinic's capacity. Further, while brief training courses are by no means adequate to prepare anyone to administer these instruments with fidelity, the alternative is to offer no formal training at all, which is unlikely to dissuade professionals interested in using the instruments from purchasing and using them on their own. In the end, professional ethics codes dictate that we practice within our competency, including only making diagnoses that we are trained to make.

Beyond diagnostic accuracy, we serve families better and more equitably when we prioritize the assessment process itself. The reality is that, though an autism diagnosis is often a significant event in a child or adult's clinical "life," there is no one treatment or service that is always needed or always effective in supporting people with autism (Thapar, Cooper, & Rutter, 2017). Thus, the autism diagnosis itself is just a start. Along with the diagnostic label, individuals with autism and their families have a right to expect to understand why they are receiving such a diagnosis, and this may vary considerably across individuals. Moreover, given the extreme heterogeneity that characterizes this population, the ASD diagnosis itself is less likely to guide treatment planning or provide meaningful information about prognosis, than information about an individual's profile of cognitive, language, and adaptive behavior skills, as well as medical and psychiatric symptoms. Thus, families also have the right to expect that recommendations for services reflect individual differences in strengths, difficulties associated with autism, and also difficulties that may be related to other commonly occurring conditions with autism, such as intellectual disability, ADHD, or depression. Each treatment recommended should be at sufficient scope and intensity to fully address the symptoms targeted.

As a field, we must find ways to hold on to best practices for diagnostic assessment: applying standardized, comprehensive methods for collecting information about current and past social communication deficits and RRBs across settings, cognitive and language abilities, adaptive behavior, and genetic, medical, psychiatric, and behavioral conditions—while also understanding how to be flexible, especially when working with special populations who were excluded from or under-represented in instrument validation samples. There are many situations in which the use of any given tool might not be possible. As such, requiring the use of specific tools, without exception, is discriminatory and damaging. More importantly, no tool or combination of tools can make up for a lack of clinical expertise in autism and related disorders. The negative consequences of misdiagnosis, resulting in part from inappropriate use of standardized instruments, have been repeatedly highlighted by clinicians, researchers, and individuals with ASD and their families as a major challenge for our field at this time. Thus, the responsibility lies with all of us to ensure that we are appropriately using any tool we choose to employ and to not allow systems to force practitioners to behave unethically or practice outside of their scope. These systems must know that *clinicians* make diagnoses, not instruments. It is the experts, who know how to apply best practice procedures for assessment beyond diagnosis, whom we are desperately lacking. This is where we must focus our resources and training efforts moving forward.

Acknowledgements

C.L. receives royalties from sales of the ADI-R and ADOS/ADOS-2. S.L.B. receives royalties from sales of the ADOS-2.

Correspondence
Somer L. Bishop, Department of Psychiatry and Behavioral Sciences and Weill Institute for Neurosciences, University of California, San Francisco, San Francisco, CA, USA; Email: somer.bishop@ucsf.edu

Key points

- When used appropriately, standardized autism diagnostic instruments may inform and enhance diagnostic decision-making and treatment planning.
- These instruments were developed to support and enhance expert clinical judgment, not to replace it.
- Use of specific instruments is not appropriate in all situations. Instrument development work should focus on creating and adapting tools, based on sound research, that can be efficiently administered to diverse patient populations by a range of providers.
- Professionals involved in the diagnostic assessment of autism must be trained in best practice methods for differential diagnosis (not just in the use of specific instruments), including how to individualize test batteries appropriate to each particular situation.

References

Bal, V.H., Maye, M., Salzman, E., Huerta, M., Pepa, L., Risi, S., & Lord, C. (2020). The adapted ADOS: A new module set for the assessment of minimally verbal adolescents and adults. *Journal of Autism and Developmental Disorders*, *50*, 719–729.

Bishop, S.L., Richler, J., & Lord, C. (2006). Association between restricted and repetitive behaviors and nonverbal IQ in children with autism spectrum disorders. *Child Neuropsychology*, *12*, 247–267.

DiLavore, P.C., Lord, C., & Rutter, M. (1995). The pre-linguistic autism diagnostic observation schedule (PLA-DOS). *Journal of Autism and Developmental Disorders*, *25*, 355–379.

Dow, D., Holbrook, A., Toolan, C., McDonald, N., Sterrett, K., Rosen, N., et al. (2022). The brief observation of symptoms of autism (BOSA): Development of a new adapted assessment measure for remote telehealth administration through COVID-19 and beyond. *Journal of Autism and Developmental Disorders*, *52*(12), 5383–5394.

Elias, R., & Lord, C. (2022). Diagnostic stability in individuals with autism spectrum disorder: Insights from a longitudinal follow-up study. *Journal of Child Psychology and Psychiatry*, *63*, 973–983.

Esbensen, A.J., Seltzer, M.M., Lam, K.S., & Bodfish, J.W. (2009). Age-related differences in restricted repetitive behaviors in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *39*, 57–66.

Evans, B. (2013). How autism became autism: The radical transformation of a central concept of child development in Britain. *History of the Human Sciences*, *26*, 3–31.

Gabrielsen, T.P., Farley, M., Speer, L., Villalobos, M., Baker, C.N., & Miller, J. (2015). Identifying autism in a brief observation. *Pediatrics*, *135*, e330–e338.

Gillberg, C. (2013). Alternative diagnostic concepts. In F.R. Volkmar (Ed.), *Encyclopedia of autism spectrum disorders*. New York, NY: Springer.

Havdahl, K.A., Bal, V.H., Huerta, M., Pickles, A., Øyen, A.S., Stoltenberg, C., … & Bishop, S.L. (2016). Multidimensional influences on autism symptom measures: Implications for use in etiological research. *Journal of the American Academy of Child & Adolescent Psychiatry*, *55*, 1054–1063.

Hus, V., Bishop, S., Gotham, K., Huerta, M., & Lord, C. (2013). Factors influencing scores on the social responsiveness scale. *Journal of Child Psychology and Psychiatry*, *54*, 216–224.

Kaat, A.J., Shui, A.M., Ghods, S.S., Farmer, C.A., Esler, A.N., Thurm, A., … & Bishop, S.L. (2020). Sex differences in scores on standardized measures of autism symptoms: A multisite integrative data analysis. *Journal of Child Psychology and Psychiatry*, *62*, 97–106.

Klaiman, C., White, S., Richardson, S., McQueen, E., Walum, H., Aoki, C., … & Klin, A. (2022). Expert clinician certainty in diagnosing autism spectrum disorder in 16–30-month-olds: A multi-site trial secondary analysis. *Journal of Autism and Developmental Disorders*, 1–16. https://doi.org/10.1007/s10803-022-05812-8

Le Couteur, A. (1989). Autism diagnostic interview: A semi-structured interview for parents and caregivers of autistic persons. *Journal of Autism and Developmental Disorders*, *19*, 363–387.

Lord, C., & Bishop, S.L. (2021). Let's be clear that "autism spectrum disorder symptoms" are not always related to autism spectrum disorder. *American Journal of Psychiatry*, *178*, 680–682.

Lord, C., Elsabbagh, M., Baird, G., & Veenstra-Vanderweele, J. (2018). Autism spectrum disorder. *Lancet*, *392*, 508–520.

Lord, C., Luyster, R.J., Gotham, K., & Guthrie, W. (2012). *Autism diagnostic observation schedule, second edition (ADOS-2) manual*. Torrance, CA: Western Psychological Services.

Lord, C., Rutter, M., DiLavore, P., Risi, S., Gotham, K., & Bishop, S. (2012). *Autism diagnostic observation schedule: ADOS-2*. Torrance, CA: Western Psychological Services.

Lord, C., Rutter, M., & DiLavore, P.C. (1999). *Autism diagnostic observation schedule – Generic*. Dissertation Abstracts International Section A: Humanities and Social Sciences.

Lord, C., Rutter, M., DiLavore, P.D., & Risi, S. (2001). *Autism diagnostic observation schedule*. Los Angeles: Western Psychological Services.

Lotter, V. (1966). Epidemiology of autistic conditions in young children. *Social Psychiatry*, *1*, 124–135.

McDonnell, C.G., Bradley, C.C., Kanne, S.M., Lajonchere, C., Warren, Z., & Carpenter, L.A. (2019). When are we sure? Predictors of clinician certainty in the diagnosis of autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *49*, 1391–1401.

Phillips, H., Wright, B., Allgar, V., McConachie, H., Sweetman, J., Hargate, R., … & Le Couteur, A. (2022). Adapting and validating the autism diagnostic observation schedule version 2 for use with deaf children and young people. *Journal of Autism and Developmental Disorders*, *52*, 553–568.

Rosen, N.E., Lord, C., & Volkmar, F.R. (2021). The diagnosis of autism: From Kanner to DSM-III to DSM-5 and beyond. *Journal of Autism and Developmental Disorders*, *51*, 4253–4270.

Rosenhall, U., Nordin, V., Sandstrom, M., Ahlsen, G., & Gillberg, C. (1999). Autism and hearing loss. *Journal of Autism Developmental Disorders*, *29*, 349–357.

Rutter, M. (1966). Behavioural and cognitive characteristics. In L. Wing & J.K. Wing (Eds.), *Early childhood autism: Clinical, educational and social aspects*. Oxford: Pergamon Press.

Rutter, M., Le Couteur, A., & Lord, C. (2003). *Autism diagnostic interview-revised*. Los Angeles: Western Psychological Services, 29, 30.

Rydzewska, E., Hughes-McCormack, L.A., Gillberg, C., Henderson, A., MacIntyre, C., Rintoul, J., & Cooper, S.A. (2018). Prevalence of long-term health conditions in adults with autism: Observational study of a whole country population. *BMJ Open*, *8*, e023945.

Spain, D., Stewart, G.R., Mason, D., Robinson, J., Capp, S.J., Gillan, N., … & Happé, F. (2022). Autism diagnostic assessments with children, adolescents, and adults prior to and during the COVID-19 pandemic: A cross-sectional survey of professionals. *Frontiers in Psychiatry*, *13*, 789449.

Thapar, A., Cooper, M., & Rutter, M. (2017). Neurodevelopmental disorders. *Lancet Psychiatry*, *4*, 339–346.

Thurm, A., Halladay, A., Mandell, D., Maye, M., Ethridge, S., & Farmer, C. (2021). Making research possible: Barriers and solutions for those with ASD and ID. *Journal of Autism and Developmental Disorders*, *52*(10), 4646–4650.

Uljarević, M., Frazier, T.W., Jo, B., Billingham, W.D., Cooper, M.N., Youngstrom, E.A., … & Hardan, A.Y. (2022). Big data approach to characterize restricted and repetitive behaviors in autism. *Journal of the American Academy of Child & Adolescent Psychiatry*, *61*, 446–457.

Wagner, L., Corona, L.L., Weitlauf, A.S., Marsh, K.L., Berman, A.F., Broderick, N.A., … & Warren, Z. (2021). Use of the TELE-ASD-PEDS for autism evaluations in response to COVID-19: Preliminary outcomes and clinician acceptability. *Journal of Autism and Developmental Disorders*, *51*, 3063–3072.

Wright, B., Phillips, H., Allgar, V., Sweetman, J., Hodkinson, R., Hayward, E., … & Le Couteur, A. (2022). Adapting and validating the autism diagnostic interview-revised for use with deaf children and young people. *Autism*, *26*, 446–459.

Zwaigenbaum, L., Bishop, S., Stone, W.L., Ibanez, L., Halladay, A., Goldman, S., … & Wetherby, A. (2021). Rethinking autism spectrum disorder assessment for children during COVID-19 and beyond. *Autism Research*, *4*, 2251–2259.

Accepted for publication: 22 March 2023